



Help Me Understand Genetics

How Genes Work

Reprinted from MedlinePlus Genetics

U.S. National Library of Medicine

National Institutes of Health

Department of Health & Human Services

Table of Contents

1	What are proteins and what do they do?	1
2	How do genes direct the production of proteins?	5
3	Can genes be turned on and off in cells?	7
4	What is epigenetics?	8
5	How do cells divide?	10
6	How do genes control the growth and division of cells?	12
7	How do geneticists indicate the location of a gene?	16

How Genes Work

1 What are proteins and what do they do?

Proteins are large, complex molecules that play many critical roles in the body. They do most of the work in cells and are required for the structure, function, and regulation of the body's tissues and organs.

Proteins are made up of hundreds or thousands of smaller units called amino acids, which are attached to one another in long chains. There are 20 different types of amino acids that can be combined to make a protein. The sequence of amino acids determines each protein's unique 3-dimensional structure and its specific function. Amino acids are coded by combinations of three DNA building blocks (nucleotides), determined by the sequence of genes.

Proteins can be described according to their large range of functions in the body, listed in alphabetical order:

Antibody. Antibodies bind to specific foreign particles, such as viruses and bacteria, to help protect the body.

Example: Immunoglobulin G (IgG) (Figure 1)

Enzyme. Enzymes carry out almost all of the thousands of chemical reactions that take place in cells. They also assist with the formation of new molecules by reading the genetic information stored in DNA.

Example: Phenylalanine hydroxylase (Figure 2)

Messenger. Messenger proteins, such as some types of hormones, transmit signals to coordinate biological processes between different cells, tissues, and organs.

Example: Growth hormone (Figure 3)

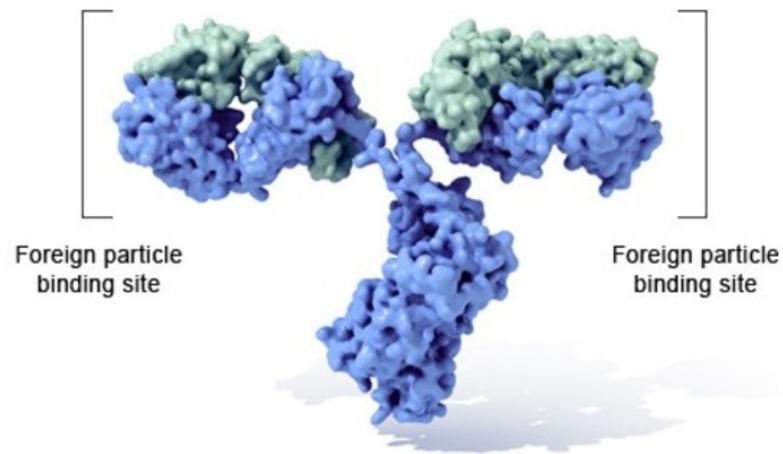
Structural component. These proteins provide structure and support for cells. On a larger scale, they also allow the body to move.

Example: Actin (Figure 4)

Transport/storage. These proteins bind and carry atoms and small molecules within cells and throughout the body.

Example: Ferritin (Figure 5)

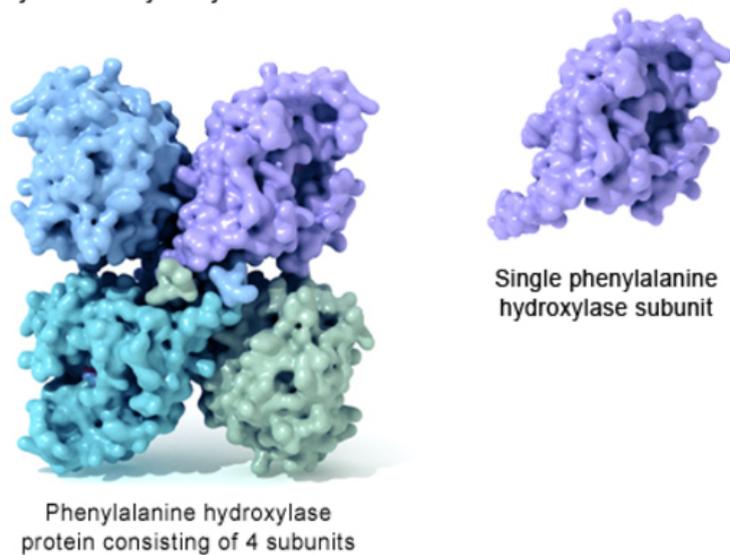
Immunoglobulin G (IgG)



U.S. National Library of Medicine

FIGURE 1: The structure of immunoglobulin G, including foreign particle binding sites.

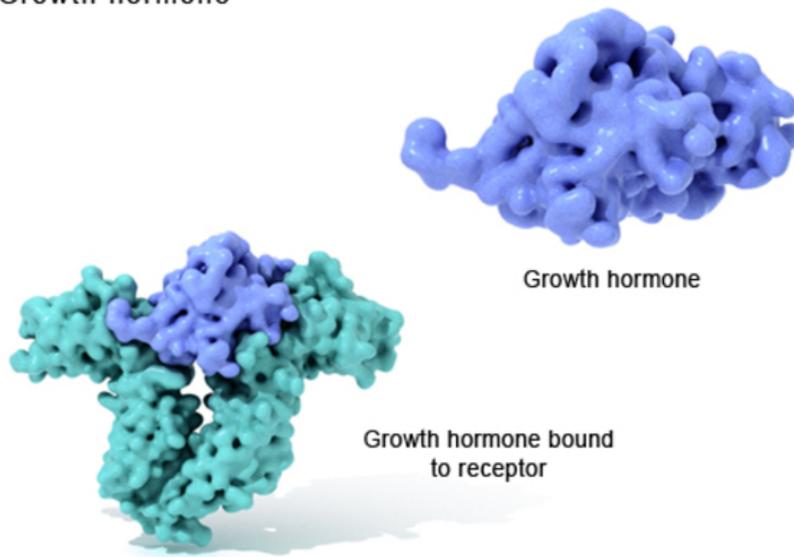
Phenylalanine hydroxylase



U.S. National Library of Medicine

FIGURE 2: The phenylalanine hydroxylase protein consists of four subunits.

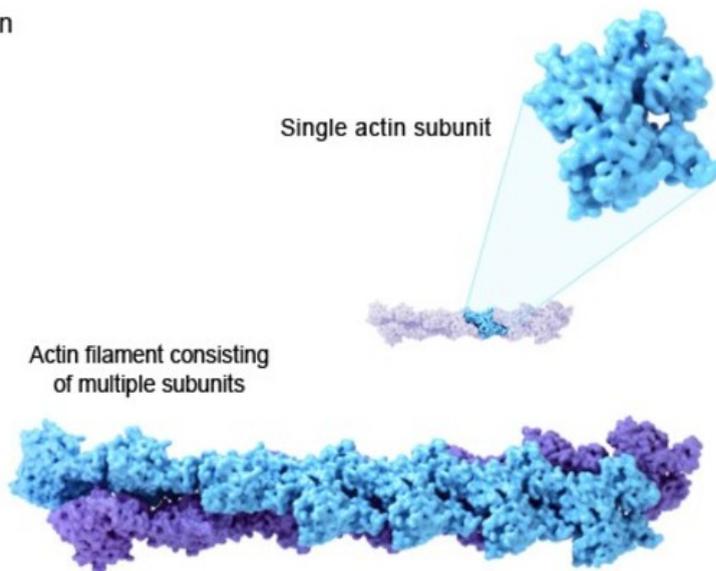
Growth hormone



U.S. National Library of Medicine

FIGURE 3: The structure of growth hormone, and growth hormone bound to its receptor.

Actin



U.S. National Library of Medicine

FIGURE 4: An actin filament consisting of multiple subunits.

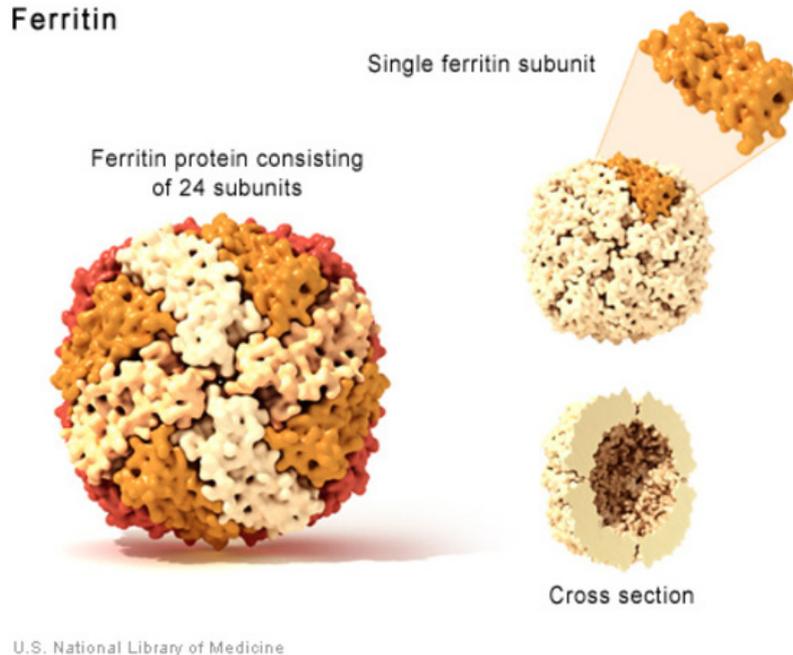


FIGURE 5: Ferritin protein consisting of 24 subunits.

For more information about proteins and their functions:

Arizona State University's "Ask a Biologist" discusses the different kinds of proteins (<http://askabiologist.asu.edu/venom/what-are-proteins>) and what they do.

The textbook *Molecular Biology of the Cell* (4th edition, 2002), from the NCBI Bookshelf, offers a detailed introduction to protein function (<https://www.ncbi.nlm.nih.gov/books/NBK26911/>).

2 How do genes direct the production of proteins?

Most genes contain the information needed to make functional molecules called proteins. (A few genes produce regulatory molecules that help the cell assemble proteins.) The journey from gene to protein is complex and tightly controlled within each cell. It consists of two major steps: transcription and translation. Together, transcription and translation are known as gene expression.

During the process of transcription, the information stored in a gene's DNA is passed to a similar molecule called RNA (ribonucleic acid) in the cell nucleus. Both RNA and DNA are made up of a chain of building blocks called nucleotides, but they have slightly different chemical properties. The type of RNA that contains the information for making a protein is called messenger RNA (mRNA) because it carries the information, or message, from the DNA out of the nucleus into the cytoplasm.

Translation, the second step in getting from a gene to a protein, takes place in the cytoplasm. The mRNA interacts with a specialized complex called a ribosome, which "reads" the sequence of mRNA nucleotides. Each sequence of three nucleotides, called a codon, usually codes for one particular amino acid. (Amino acids are the building blocks of proteins.) A type of RNA called transfer RNA (tRNA) assembles the protein, one amino acid at a time. Protein assembly continues until the ribosome encounters a "stop" codon (a sequence of three nucleotides that does not code for an amino acid).

The flow of information from DNA to RNA to proteins is one of the fundamental principles of molecular biology. It is so important that it is sometimes called the "central dogma."

Through the processes of transcription and translation, information from genes is used to make proteins. (Figure 6)

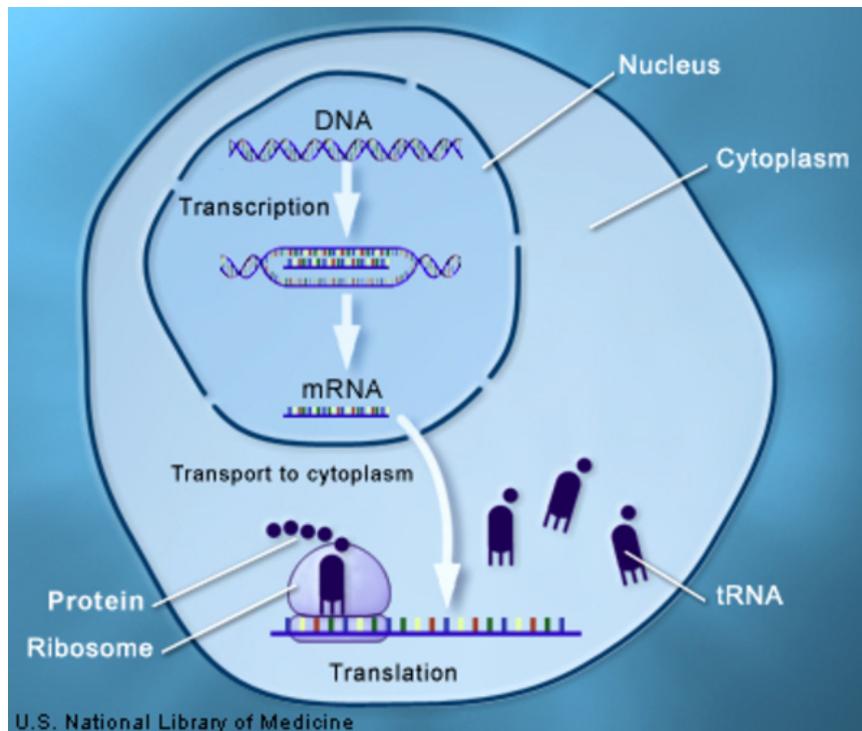


FIGURE 6: DNA is transcribed into mRNA, which is transported into the cell cytoplasm and translated into a protein.

For more information about making proteins:

The Genetic Science Learning Center at the University of Utah offers an interactive introduction to transcription and translation (<https://learn.genetics.utah.edu/content/basics/dna>).

The New Genetics (<https://www.nigms.nih.gov/education/Booklets/the-new-genetics/Pages/Home.aspx>), a publication of the National Institute of General Medical Sciences, includes discussions of transcription and translation.

Biointeractive from the Howard Hughes Medical Institute illustrates the stages in the flow of information from DNA to RNA to protein. (<https://www.biointeractive.org/classroom-resources/central-dogma-and-genetic-medicine>) This tool also gives examples of how modern technologies that target the different stages are used to treat genetic diseases.

3 Can genes be turned on and off in cells?

Each cell expresses, or turns on, only a fraction of its genes at any given time. The rest of the genes are repressed, or turned off. The process of turning genes on and off is known as gene regulation. Gene regulation is an important part of normal development. Genes are turned on and off in different patterns during development to make a brain cell look and act different from a liver cell or a muscle cell, for example. Gene regulation also allows cells to react quickly to changes in their environments. Although we know that the regulation of genes is critical for life, this complex process is not yet fully understood.

Gene regulation can occur at any point during gene expression, but most commonly occurs at the level of transcription (when the information in a gene's DNA is passed to mRNA). Signals from the environment or from other cells activate proteins called transcription factors. These proteins bind to regulatory regions of a gene and increase or decrease the level of transcription. By controlling the level of transcription, this process can determine when and how much protein product is made by a gene.

For more information about gene regulation:

The National Human Genome Research Institute provides a definition of gene regulation (<https://www.genome.gov/genetics-glossary/Gene-Regulation>) in their Talking Glossary of Genetic Terms.

The Genetic Science Learning Center at the University of Utah offers an explanation of gene expression as it relates to disease risk (<https://learn.genetics.utah.edu/content/science/expression/>).

Additional information about gene expression (<https://www.yourgenome.org/facts/what-is-gene-expression>) is available from yourgenome.org, a service of the Wellcome Trust.

The Khan Academy has an educational unit on gene regulation (<https://www.khanacademy.org/science/biology/gene-regulation>), including videos about gene regulation in bacteria and eukaryotes.

4 What is epigenetics?

Epigenetics is the study of how cells control gene activity without changing the DNA sequence. "Epi-" means on or above in Greek, and "epigenetic" describes factors beyond the genetic code. Epigenetic changes are modifications to DNA that regulate whether genes are turned on or off. These modifications are attached to DNA and do not change the sequence of DNA building blocks. Within the complete set of DNA in a cell (genome), all of the modifications that regulate the activity (expression) of the genes is known as the epigenome.

Because epigenetic changes help determine whether genes are turned on or off, they influence the production of proteins in cells. This regulation helps ensure that each cell produces only proteins that are necessary for its function. For example, proteins that promote bone growth are not produced in muscle cells. Patterns of epigenetic modification vary among individuals, in different tissues within an individual, and even in different cells within a tissue. Environmental influences, such as a person's diet and exposure to pollutants, can impact the epigenome. Epigenetic modifications can be maintained from cell to cell as cells divide and, in some cases, can be inherited through the generations.

A common type of epigenetic modification is called DNA methylation. DNA methylation involves the attachment of small chemical groups called methyl groups (each consisting of one carbon atom and three hydrogen atoms) to DNA building blocks. When methyl groups are present on a gene, that gene is turned off or silenced, and no protein is produced from that gene.

Another common epigenetic change is histone modification. Histones are structural proteins in the cell nucleus. DNA wraps around histones, giving chromosomes their shape. Histones can be modified by the addition or removal of chemical groups, such as methyl groups or acetyl groups (each consisting of two carbon, three hydrogen, and one oxygen atoms). The chemical groups influence how tightly the DNA is wrapped around histones, which affects whether a gene can be turned on or off.

Errors in the epigenetic process, such as modification of the wrong gene or failure to add a chemical group to a particular gene or histone, can lead to abnormal gene activity or inactivity. Altered gene activity, including that caused by epigenetic errors, is a common cause of genetic disorders. Conditions such as cancers, metabolic disorders, and degenerative disorders have been found to be related to epigenetic errors.

Scientists continue to explore the relationship between the genome and the chemical compounds that modify it. In particular, they are studying the effects that epigenetic modifications and errors have on gene function, protein production, and human health.

For more information about the epigenome:

The National Institutes of Health (NIH) offers the NIH Roadmap Epigenomics Project (<http://www.roadmapepigenomics.org/>), which provides epigenome maps of a variety of cells to begin to assess the relationship between epigenomics and human disease.

Human Epigenome Toolset (http://www.genboree.org/site/epigenomics_toolset) from Baylor College of Medicine allows for comparison of the epigenomes of many species and cell types.

Ongoing research is being done with the International Human Epigenome Consortium (<http://ihec-epigenomes.org/>).

The University of Utah provides an interactive epigenetics tutorial (<https://learn.genetics.utah.edu/content/epigenetics/>).

The National Human Genome Research Institute provides a fact sheet (<https://www.genome.gov/about-genomics/fact-sheets/Epigenomics-Fact-Sheet>) on Epigenomics.

Many tools for understanding epigenomics are available through the NIH Common Fund Epigenomics Project (<https://commonfund.nih.gov/epigenomics/>).

5 How do cells divide?

There are two types of cell division: mitosis and meiosis. Most of the time when people refer to “cell division,” they mean mitosis, the process of making new body cells. Meiosis is the type of cell division that creates egg and sperm cells.

Mitosis is a fundamental process for life. During mitosis, a cell duplicates all of its contents, including its chromosomes, and splits to form two identical daughter cells. Because this process is so critical, the steps of mitosis are carefully controlled by certain genes. When mitosis is not regulated correctly, health problems such as cancer can result.

The other type of cell division, meiosis, ensures that humans have the same number of chromosomes in each generation. It is a two-step process that reduces the chromosome number by half—from 46 to 23—to form sperm and egg cells. When the sperm and egg cells unite at conception, each contributes 23 chromosomes so the resulting embryo will have the usual 46. Meiosis also allows genetic variation through a process of gene shuffling while the cells are dividing.

Mitosis and meiosis, the two types of cell division. (Figure 7)

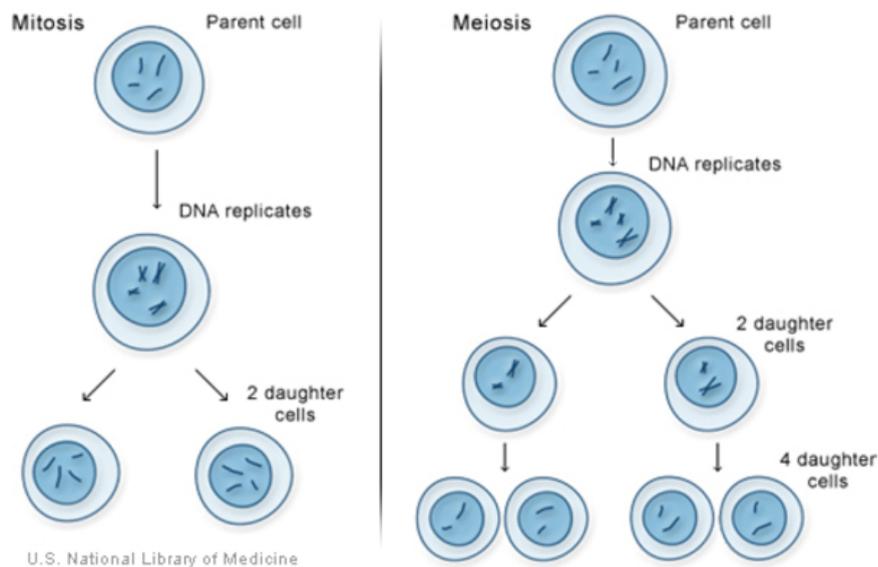


FIGURE 7: The cell division processes of mitosis and meiosis.

For more information about cell division:

The National Human Genome Research Institute's Talking Glossary provides

information about mitosis (<https://www.genome.gov/genetics-glossary/Mitosis>) and meiosis (<https://www.genome.gov/genetics-glossary/Meiosis>).

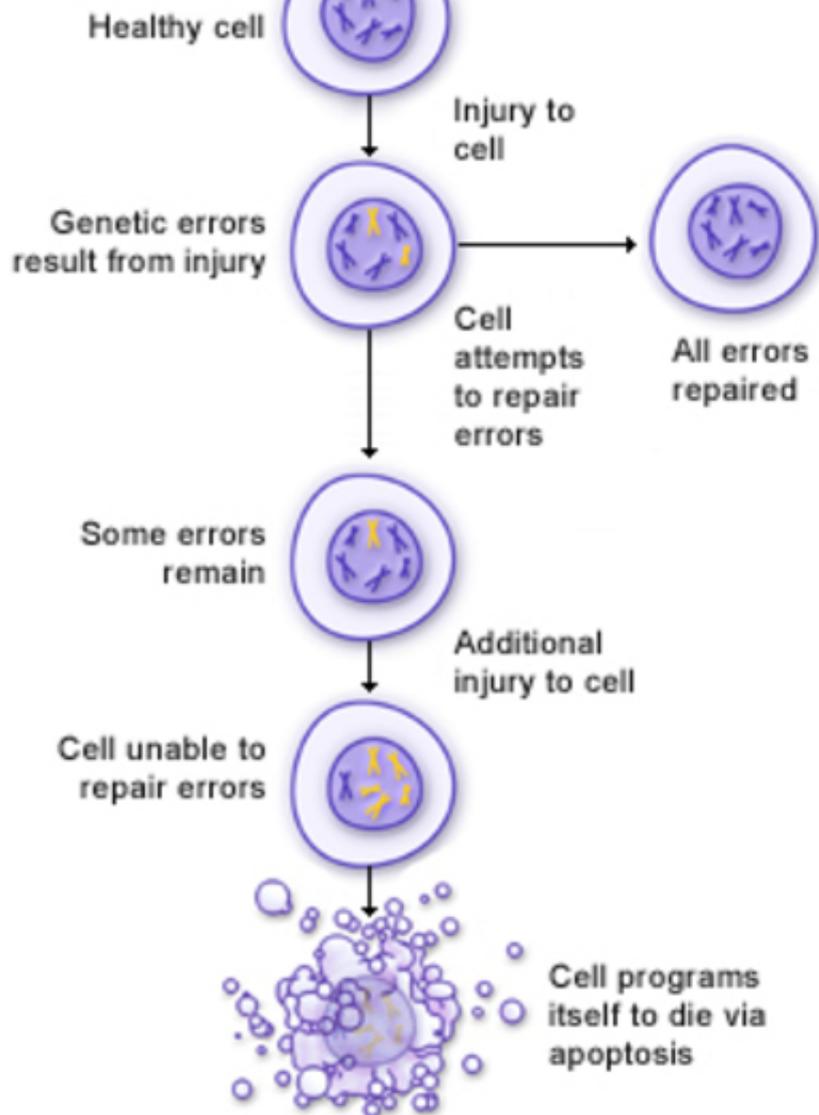
Yourgenome from the Wellcome Trust outlines the similarities and differences between mitosis and meiosis (<https://www.yourgenome.org/facts/mitosis-versus-meiosis>).

6 How do genes control the growth and division of cells?

A variety of genes are involved in the control of cell growth and division. The cell replicates itself in an organized, step-by-step fashion known as the cell cycle. Tight regulation of this process ensures that a dividing cell's DNA is copied properly, any errors in the DNA are repaired, and each daughter cell receives a full set of chromosomes. The cell cycle has checkpoints (also called restriction points), which allow certain genes to check for problems and halt the cycle for repairs if something goes wrong.

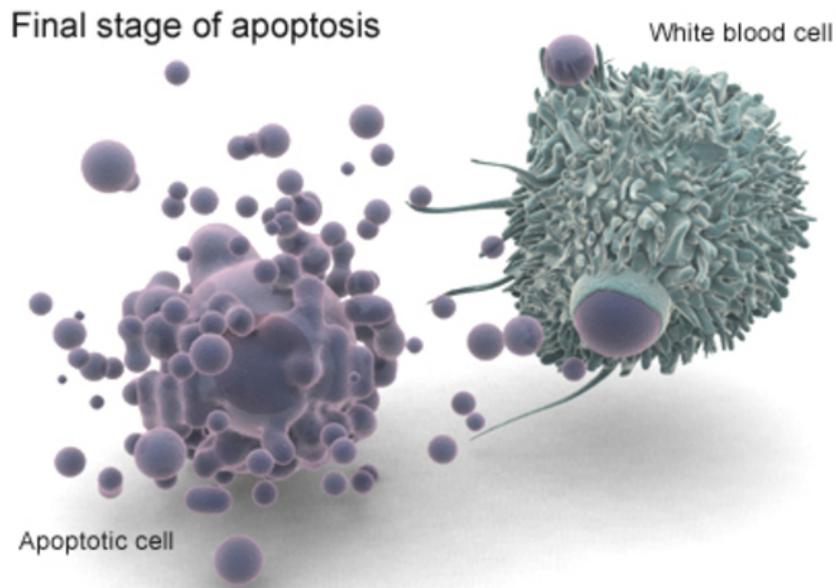
If a cell has an error in its DNA that cannot be repaired, it may undergo self-destruction (apoptosis (Figure 8)). Apoptosis is a common process throughout life that helps the body get rid of cells that no longer work or that it doesn't need. Cells that undergo apoptosis break apart (Figure 9) and are recycled by a type of white blood cell called a macrophage. Apoptosis protects the body by removing genetically damaged cells that could lead to cancer, and it plays an important role in the development of the embryo and the maintenance of adult tissues.

Apoptosis



U.S. National Library of Medicine

FIGURE 8: If a cell is unable to repair errors, it programs itself to die via apoptosis.

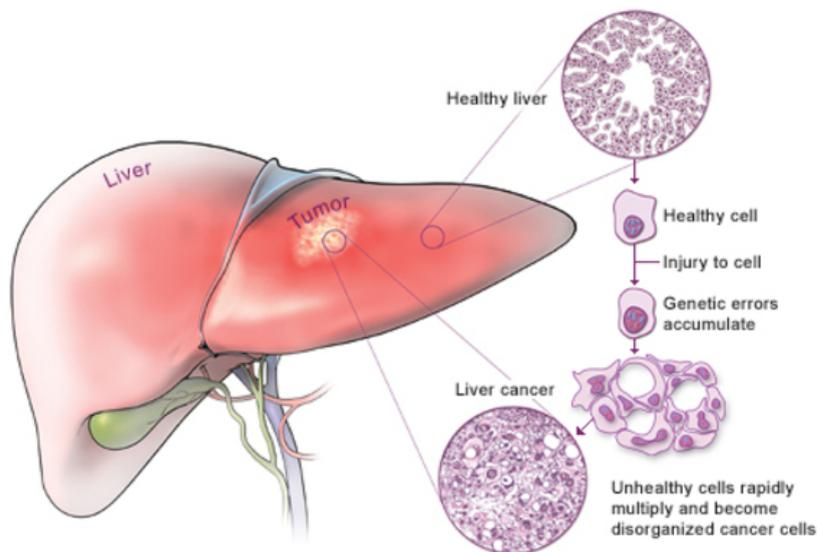


U.S. National Library of Medicine

FIGURE 9: A white blood cell and an apoptotic cell.

Disruption of normal regulation of the cell cycle can lead to diseases such as cancer. When the cell cycle proceeds without control, cells can divide without order and accumulate genetic errors that can lead to a cancerous tumor (Figure 10).

Genetic mutation and cancer development



U.S. National Library of Medicine

FIGURE 10: A tumor in the liver; inset shows the progression from a normal cell to one in which genetic changes accumulate, leading to rapid

and disorganized cell division.

For more information about cell growth and division:

The National Human Genome Research Institute's Talking Glossary provides information about the cell cycle (<https://www.genome.gov/genetics-glossary/Cell-Cycle>).

The National Cancer Institute's fact sheet What is Cancer? (<https://www.cancer.gov/about-cancer/understanding/what-is-cancer>) explains the growth of cancerous tumors.

7 How do geneticists indicate the location of a gene?

Geneticists use maps to describe the location of a particular gene on a chromosome. One type of map uses the cytogenetic location to describe a gene's position. The cytogenetic location is based on a distinctive pattern of bands created when chromosomes are stained with certain chemicals. Another type of map uses the molecular location, which is a precise description of a gene's position on a chromosome. The molecular location is based on the sequence of DNA building blocks (nucleotides) that make up the chromosome.

Cytogenetic location

Geneticists use a standardized way of describing a gene's cytogenetic location. In most cases, the location describes the position of a particular band on a stained chromosome:

17q12

It can also be written as a range of bands, if less is known about the exact location:

17q12-q21

The combination of numbers and letters provide a gene's "address" on a chromosome. This address is made up of several parts:

- The chromosome on which the gene can be found. The first number or letter used to describe a gene's location represents the chromosome. Chromosomes 1 through 22 (the autosomes) are designated by their chromosome number. The sex chromosomes are designated by X or Y.
- The arm of the chromosome. Each chromosome is divided into two sections (arms) based on the location of a narrowing (constriction) called the centromere. By convention, the shorter arm is called p, and the longer arm is called q. The chromosome arm is the second part of the gene's address. For example, 5q is the long arm of chromosome 5, and Xp is the short arm of the X chromosome.
- The position of the gene on the p or q arm. The position of a gene is based on a distinctive pattern of light and dark bands that appear when the chromosome is stained in a certain way. The position is usually designated by two digits (representing a region and a band), which are sometimes followed by a decimal point and one or more additional digits (representing sub-bands within a light or dark area). The number indicating the gene position increases with distance from the centromere. For example: 14q21 represents position 21 on the long arm of chromosome 14. 14q21 is closer to the centromere than 14q22.

Sometimes, the abbreviations "cen" or "ter" are also used to describe a gene's cytogenetic location. "Cen" indicates that the gene is very close to the centromere. For example, 16pcen refers to the short arm of chromosome 16 near the centromere. "Ter" stands for terminus, which indicates that the gene is very close to the end of the p or q

arm. For example, 14qter refers to the tip of the long arm, or the very end, of chromosome 14.

The CFTR gene is located on the long arm of chromosome 7 at position 7q31.2. (Figure 11)

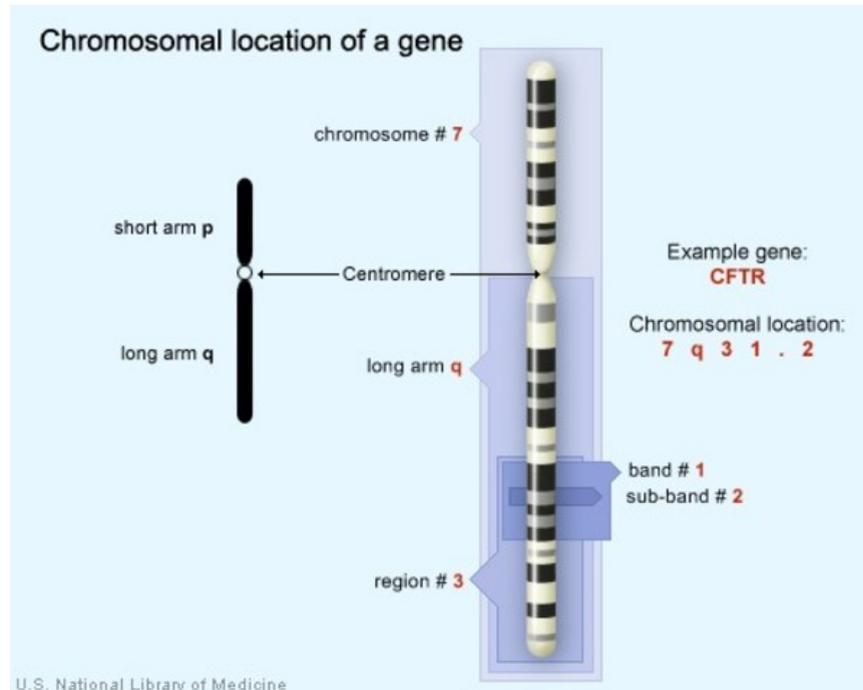


FIGURE 11: The CFTR gene is located on the long (q) arm of chromosome 7 at position 31.2.

Molecular location

The Human Genome Project, an international research effort completed in 2003, determined the sequence of nucleotides for each human chromosome. This sequence information allows researchers to provide a more specific address than the cytogenetic location for many genes. A gene's molecular address pinpoints the location of that gene in terms of nucleotides. It describes the gene's precise position on a chromosome and indicates the size of the gene. Knowing the molecular location also allows researchers to determine exactly how far a gene is from other genes on the same chromosome.

Different groups of researchers often present slightly different values for a gene's molecular location. Researchers interpret the sequence of the human genome using a variety of methods, which can result in small differences in a gene's molecular address.

For more information on genetic mapping:

The National Human Genome Research Institute explains how researchers create a genetic map (<https://www.genome.gov/about-genomics/fact-sheets/Genetic-Mapping-Fact-Sheet>).

Information about assembling and annotating the genome (<https://www.ncbi.nlm.nih.gov/guide/genomes-maps/>) is available from NCBI.